



Howard, P., Liversedge, S. P., & Benson, V. (2017). Investigating the Use of World Knowledge During On-line Comprehension in Adults with Autism Spectrum Disorder. *Journal of Autism and Developmental Disorders*, 47(7), 2039–2053. <https://doi.org/10.1007/s10803-017-3129-x>

Peer reviewed version

Link to published version (if available):  
[10.1007/s10803-017-3129-x](https://doi.org/10.1007/s10803-017-3129-x)

[Link to publication record in Explore Bristol Research](#)  
PDF-document

This is the author accepted manuscript (AAM). The final published version (version of record) is available online via Springer Nature at <https://link.springer.com/article/10.1007%2Fs10803-017-3129-x>. Please refer to any applicable terms of use of the publisher.

## University of Bristol - Explore Bristol Research

### General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available:  
<http://www.bristol.ac.uk/red/research-policy/pure/user-guides/ebr-terms/>

Investigating the Use of World Knowledge During On-line Comprehension in Adults with  
Autism Spectrum Disorder.

Philippa L. Howard\*

Valerie Benson

Simon P. Liversedge

University of Southampton

**Author accepted version. The final publication is available at Springer via  
<http://dx.doi.org/10.1007/s10803-017-3129-x>**

\*Corresponding author: [plh1g11@soton.ac.uk](mailto:plh1g11@soton.ac.uk)

Department of Psychology

Shackleton Building (B44)

University of Southampton

Highfield Campus

Southampton

SO17 1BJ

### Abstract

The on-line use of world knowledge during reading was examined in adults with autism spectrum disorder (ASD). Both ASD and typically developed (TD) adults read sentences that included plausible, implausible and anomalous thematic relations, as their eye movements were monitored. No group differences in the speed of detection of the anomalous violations were found, but the ASD group showed a delay in detection of implausible thematic relations. These findings suggest that there are subtle differences in the speed of world knowledge processing during reading in ASD.

**Keywords:** Autism spectrum disorder, reading, sentence processing, semantics, world knowledge, plausibility.

Autism Spectrum Disorder (ASD) is a neurodevelopmental condition characterised by communication/social interaction difficulties and restricted and repetitive behaviour (American Psychiatric Association, 2013). There is a large body of literature to suggest that this unique behavioural phenotype is underpinned by cognitive processing differences (Frith, 2012).

Language is one area where cognitive differences manifest in ASD (Tager-Flusberg, 1981) and consistent with this, reading ability is often found to be atypical. The population of individuals diagnosed with ASD is heterogeneous and as a consequence, reading ability is highly variable (Nation, Clarke, Wright, & Williams, 2006). However, there is a general finding that individuals with ASD and no known learning difficulties or evidence of additional language impairment perform comparably to typically developing (TD) participants on low-level linguistic tasks, such as word identification (Howard, Liversedge & Benson, 2017; Huemer & Mann, 2010; Mayes & Calhoun, 2006; Minshew, Goldstein, & Siegel, 1995; Saldaña, Carreiras, & Frith, 2009; Frith & Snowling, 1983, but see also Åsberg & Sandberg, 2012; Nation et al., 2006; White et al., 2006, who found subgroups of children with ASD to have less accurate word reading). Conversely, performance accuracy for tasks that require higher order linguistic processing, such as text comprehension and inferencing, is generally reported to be less accurate in comparison to TD controls (Brown, Oram-Cardy, & Johnson, 2013; Huemer & Mann, 2010; Jolliffe & Baron-Cohen, 1999; Nation et al., 2006; Newman et al., 2007; Minshew et al., 1995; Jones et al., 2009, but see also Åsberg, Kopp, Berg-Kelly, & Gillberg, 2010; Saldaña & Frith, 2007 who found no differences, and Lucas & Norbury, 2014, 2015; Norbury & Nation, 2011 who found performance differences in ASD to be associated with additional language impairment). It is these group differences in performance for higher order linguistic tasks that are of interest in the present work, because

they often cannot be attributed to poor basic reading skill. Hence, it is possible that performance difficulties are associated with ASD specific cognitive processing differences.

The Weak Central Coherence Theory (WCC) proposes that individuals with ASD have a domain general local processing bias (Frith, 1989; Frith & Happé, 1994; Happé & Frith, 2006). Some researchers have suggested that the associated lack of a global processing bias and consequential integration difficulties, may underpin reading difficulties in ASD. For example, readers with ASD are found to be less accurate at modulating their pronunciation of a homograph (word with two spellings, one meaning e.g., *tear* meaning cry or rip) based upon the (global) sentence context (Happé, 1997; Jolliffe & Baron-Cohen, 1999; López & Leekam, 2003; Frith & Snowling, 1983, but see also Snowling & Frith, 1986, where performance was modulated by verbal ability). However, concerns have been raised as to the methodology employed, and the assumptions and conclusions that have been made about the cognitive processes a participant has to engage in to be successful at this task (Brock & Bzishvili, 2013; Brock & Caruana, 2014). Furthermore, participants with ASD who do not have language impairment, are successful at modulating their pronunciation of a homograph and are sensitive to contextual linguistic information when other paradigms are adopted, such as semantic priming (Hala, Pexman, & Glenwright, 2007; Henderson, Clarke, & Snowling, 2011; Norbury, 2005), eye movements in a visual world (Brock, Norbury, Einav, & Nation, 2008; Hahn, Snedeker, & Rabagliati, 2015) and eye movements and reading (Au-Yeung, Kaakinen, Liversegde, & Benson, 2015, also see Caruana & Brock, 2014 for evidence of online contextual processing during reading in a group of adults with high levels of self reported autistic traits). Therefore, it would seem that difficulties in the construction of a mental representation of text (integration), as is predicted by WCC theory, is not a driving force behind difficulties with higher order linguistic tasks in ASD.

An alternative theory that has attempted to explain the cognitive differences in ASD is the Theory of Complex Information Processing (CIP; Minshew & Goldstein, 1998; Minshew, Goldstein & Siegel, 1997; Williams, Goldstein & Minshew, 2006). The CIP theory is a description of the behavioural and cognitive outcomes that you would expect from the Under-Connectivity Hypothesis, which proposes that ASD is a result of under-connectivity between neocortical brain areas (Just, Cherkassky, Keller, Kana, & Minshew, 2007; Just, Cherkassky, Keller, & Minshew, 2004; Minshew, Williams, & McFadden, 2008). The CIP theory posits that individuals with ASD have intact performance for ‘simple’ tasks, defined in the context of linguistic processing as those tasks that can be completed upon the basis of explicit rules (e.g., syntax) or information deducible from the stimuli (e.g., word meaning), whereas performance differences will be present for ‘complex’ tasks, defined as tasks that require processing beyond what is explicitly stated within a text (e.g., Minshew & Goldstein, 1998; Minshew et al. 1997; Williams et al. 2006). Consistent with the CIP’s predictions, research using ‘simple’ linguistic tasks tends to show similar performance for TD and ASD readers, whereas research using ‘complex’ tasks show differences (e.g., tasks relying on word identification vs. tasks requiring text comprehension and inferencing in Minshew et al., 1995).

Therefore, the CIP theory posits that processing will differ between TD and ASD groups when a reader is required to use knowledge that is not explicitly provided in the text. But, for comprehension to succeed, it is often the case that a reader must infer such information on the basis of schematic knowledge of the world that is gained and developed through life experience, and stored in long-term memory (e.g., episodic, procedural, semantic, Gernsbacher, 1991; Kintsch, 1988; Zwaan, Langston & Graesser, 1995).

1. John got distracted whilst running his bath. He sighed as he mopped up the sodden bathroom floor. Why did John have to mop the bathroom floor?

For example, in example 1 above, most readers would confidently answer the comprehension question with a response such as “the bath overflowed”. However, this information is not explicitly provided in the preceding sentence; the reader must infer that the bath overflowed and flooded the bathroom floor, based upon their knowledge of baths, taps, water and distraction (world knowledge). Note that the relative ease in answering this question reflects the automaticity of activation of this information during natural reading.

As demonstrated above, the incremental evaluation of world knowledge is fundamental for inferential processing and the comprehension of text in order for local and global coherence to be gained. If readers with ASD do have deficits in these processes, this would significantly impact upon their understanding of text and may contribute to the commonly reported performance differences in tasks that require a reader to engage in such processes.

There is evidence of performance difficulties in ASD during reading tasks that require the use of world knowledge. For example, there have been numerous reports of participants with ASD performing less accurately than controls when they are asked to answer comprehension questions about a story they have read or heard that requires inferential work (Bodner, Engelhardt, Minshew & Williams, 2015; Jolliffe & Baron-Cohen, 1999, 2000; Dennis, Lazenby & Lockyer, 2001; Minshew et al., 1995; Norbury & Bishop, 2002; Norbury & Nation; 2011, see also Lucas & Norbury, 2014 who found performance in children to be associated with verbal working memory and vocabulary). In addition, Norbury and Bishop (2002) identified that children with ASD were more likely than children with specific or pragmatic language impairment to have difficulties making inferences and ASD symptomology has been found to account for unique variance associated with inferential skill (Bodner et al., 2015; Norbury & Bishop, 2002; Norbury & Nation, 2011). Inferencing requires the activation and evaluation of relevant world knowledge that is then incorporated

into the reader's mental representation of a described event (Graesser, Singer, & Trabasso, 1994; McKoon & Ratcliff, 1992). It has been concluded in the majority of the studies cited above that a deficit in construction of a discourse representation (integration difficulties), is likely to contribute to poor performance accuracy, as predicted by the WCC theory. However, it is also possible that the use and evaluation of world knowledge in ASD may be the underlying cause of such difficulties.

In an attempt to evaluate whether readers use world knowledge during reading, Saldaña and Frith (2007) tasked participants with and without ASD to read two sentence vignettes that required a bridging inference in order for successful comprehension to be attained, followed by a comprehension question that was or was not related to the inference. Both groups read questions that were related to the inference faster than they read questions that were not. Saldaña and Frith (2007) concluded that the lack of difference between the TD and ASD groups was evidence of intact on-line use of world knowledge. However, question-reading time that follows the computation of an inference does not necessarily reflect the moment-to-moment cognitive processes that occur during normal reading, and therefore it is possible that this approach was not sufficiently sensitive to allow detection of on-line processing differences between ASD and TD groups.

A study conducted by Sansosti, Was, Rawson, and Remaklus (2013) has attempted to address this issue by replicating Saldaña and Frith's (2007) experiment. However, in this study, eye movements were recorded as participants processed sentences, because there is a strong relationship between when and where readers make fixations and on-line cognitive processes readers engage in to comprehend text (Liversedge & Findlay, 2000). Sansosti et al. (2013) reported global measures of reading behaviour in their study. Note however, that global eye movement measures are calculated based upon entire vignette reading times; they do not offer the opportunity to establish the precise point in sentence processing at which



participants first experienced difficulty. Sansosti et al. (2013) did, however, report that the ASD group made significantly more and longer fixations and an increased number of regressions in comparison to the TD group and concluded that this was evidence of an integration deficit when a bridging inference was required for the construction of a coherent discourse representation. However, as already noted, by only examining global eye movement measures it is not possible to explore the time course of such processing during normal reading. Local reading time measures associated with specific words and critical regions in carefully constructed experimental sentences are necessary to form conclusions about on-line processing during reading (Rayner, 1998; 2009).

A recent study required TD and ASD participants to read garden path sentences that contained an ambiguous prepositional phrase that could either be attached high to the verb (e.g., 2a, target word italicized) or low as a modifier to the noun phrase (e.g., 2b, Howard et al., 2016), as their eye movements were monitored.

2a. Charlie demolished the dilapidated house with a huge *crane* last year.

2b. Charlie demolished the dilapidated house with a huge *fence* last year.

Typical readers show a preference to attach ambiguous prepositional phrases high (e.g. Rayner, Carlson & Frazier, 1983, but see also Taraban & McClelland, 1988). Therefore, when encountering sentences in which such a prepositional phrase attachment preference results in a semantic anomaly that conflicts with world knowledge (e.g., in 2b, a fence is not a tool and therefore not something that could be used to demolish a house), disruption to reading occurs as a result of readers having to re-evaluate their initial structural interpretation of the sentence. This disruption to reading results in increased fixation times upon the disambiguating target and increased regressions back to reread previous portions of the text. Howard et al. (2016) found adults with ASD to show an onset and magnitude of reading disruption when reading low attached sentences that was very comparable to TD controls.

This suggests that not only did readers with ASD adopt a high attachment preference, but they also appeared to be as efficient as TD readers in the use of world knowledge on-line to detect an initial syntactic misanalysis.

The aim of the current experiment was to further examine the on-line evaluation of world knowledge during natural reading in ASD. To achieve this, we recorded eye movements as participants read sentences containing semantic oddities differing in the severity with which they violate world knowledge. This approach has been employed to investigate the immediacy with which world knowledge is activated and used in skilled adult readers (Rayner, Warren, Juhasz, & Liversedge, 2004).

3a. John used a knife to chop the large *carrots* for dinner last night. (Plausible)

3b. John used an axe to chop the large *carrots* for dinner last night. (Implausible)

3c. John used a pump to inflate the large *carrots* for dinner last night. (Anomalous)

Consider sentences 3b and 3c above. In order to recognise that the events described in these sentences are odd or unusual, each event must be evaluated against what is known to be true about the world, for example, knowledge about carrots and how they are normally prepared for a meal. When such sentences are understood to mean something that is inconsistent with such knowledge, the detection of that inconsistency has been demonstrated to result in disruption to eye movement behaviour during reading. The immediacy and the nature of such disruption provides insight into the time course of the use of world knowledge during reading. Rayner et al. (2004) demonstrated this by recording the eye movements of a TD group of participants as they read sentences that described events that were plausible (control e.g., 3a), implausible (possible but unlikely e.g., 3b) or anomalous (impossible e.g., 3c). In each of the sentences the target word is *carrots*, and it is at this word in the implausible and anomalous sentences that the semantic oddity first becomes apparent to the reader.

Specifically, the anomalous sentences include a verb argument violation (i.e., a carrot cannot

be inflated), whereas in the implausible sentences there is a mismatch in the co-occurrence of two quite reasonable verb arguments (i.e., an axe can be quite reasonably used to chop things, and carrots can be quite reasonably chopped, but the use of an axe to chop carrots, whilst possible, is unlikely). Rayner et al. (2004) found that the detection of an anomaly was almost immediate, with readers having significantly longer gaze durations (the duration of time spent fixating a word until the eyes leave that word to the left or right) on the target word in comparison to the control sentences. Implausibilities were also shown to be disruptive to reading, however disruption was less immediate, becoming apparent later in the eye movement record, with go past times being increased on the words that immediately followed the target (go past time sums the time from when a word is first fixated, until a fixation to the right of the word, therefore including any rereading of previous text). These effects of anomaly and implausibility on linguistic processing have been replicated in adults and children (Joseph et al., 2008) and are found to occur extremely rapidly and incrementally, with disruption to initial processing occurring even when a prior context licenses a world knowledge violation, such as fictional contexts (Warren, McConnell, & Rayner, 2008) and counterfactual statements (Ferguson & Sanford, 2008).

The disruption caused by these manipulations is thought to reflect the difficulty readers have with building a mental representation of the events when these events conflict with their knowledge of the world. There are two possible reasons for the difference in the onset of disruption for anomalous and implausible sentences. Firstly, it may be a result of the difference in the severity of the semantic oddity between the sentences, with anomalies being more severe violations than implausibilities. Secondly, there is evidence that the anomalies may be detected at an earlier stage of processing independent of world knowledge evaluation, when thematic roles are assigned, as a result of the violation of a verb's selectional restrictions (semantic rules about what can and cannot be an argument to the verb e.g.,

Warren & McConnell, 2007).

In this study we adopted the paradigm used by Rayner et al. (2004) and invited adults with and without ASD to take part. Global off-line reading times for semantically anomalous words have previously been found to be similar between TD and ASD children,(Lucas & Norbury, 2014), however, we will use the technique described above to establish whether there are any differences in the time course of world knowledge evaluation during natural reading in an adult sample of readers with ASD, in comparison to a TD group. We predict that, consistent with previous findings (Joseph et al. 2008; Rayner et al. 2004), the TD group will detect anomalies more rapidly than implausibilities and that anomalies will result in increased disruption to reading, relative to implausibilities. We also predict, based upon the hypothesis that ASD participants will be less efficient in the use of world knowledge and the assumption that both implausibilities and anomalies become apparent to the reader via world knowledge evaluation, that the detection of implausibilities and anomalies will be delayed in the ASD group, in comparison to the TD group.

## **Method**

### **Participants**

Two groups of adults were recruited (aged 18+), 24 with a clinical diagnosis of an ASD (five females), and 24 who were part of the TD control group (six females). All participants had normal or corrected to normal vision, were native English speakers and had no diagnosed reading difficulties (e.g., dyslexia). Participants with ASD were recruited through advertisement via local charitable organisations, with 21 members of the sample having received a clinical diagnosis of Asperger's syndrome, one member receiving a diagnosis of pervasive developmental disorder, and two members a diagnosis of autism. Diagnostic reports confirmed that all participants were primarily diagnosed using standard diagnostic instruments, including the Autism Diagnostic Observation Schedule (Lord, Rutter,

DiLavore, & Risi, 2001), and/or the Autism Diagnostic Interview–Revised (Lord, Rutter, & Le Couteur, 1994). Control participants were recruited from the local community via online and poster advertisement. All participants gave written informed consent and were paid for their time.

All participants were assessed for oral language difficulties by completing the sentence repetition subtest of the Clinical Evaluation of Language Fundamentals II (CELF; Semel, Wiig & Secord, 2003), which is an assessment of expressive language production and verbal working memory that is sensitive to difficulties associated with specific language impairment (e.g., Conti-Ramsden, Botting, & Faragher, 2001). All TD participants and 22 of the ASD participants scored highly, with raw scores attained falling above the highest age equivalent score available (>12.11 years). The two participants in the ASD group who scored below this cut off (both males, one who had a diagnosis of Asperger’s syndrome and one with a diagnosis of autism) were excluded from analysis to avoid any confounds associated with oral language impairment (e.g., Lucas & Norbury, 2014, 2015; Norbury & Nation, 2011; Norbury, 2005). The remaining sample of 22 ASD and 24 TD participants, did not differ on average for performance on the sentence repetition subtest of the CELF;  $t(43.73) = 0.54, p = .594$ . All participants were also in the normal range of intelligence (>80), as measured by the Weschler Abbreviated Scale of Intelligence (WASI; Wechsler, 1999) and the two groups did not differ in verbal IQ;  $t(41.06) = 0.24, p = .815$ , performance IQ;  $t(42.99) = 1.13, p = .266$ , or full scale IQ;  $t(42.53) = 0.62, p = .538$ . The Secondary Version of the York Assessment of Reading Comprehension (Snowling et al. 2010) was administered to all participants and raw scores from single word reading revealed no group differences in word identification accuracy;  $t(43.66) = 0.86, p = .396$ , or passage comprehension accuracy;  $t(42.42) = 1.38, p = .174$ . The ASD group did however have significantly higher levels of self-reported autistic traits in comparison to the TD group as measured by the Autistic Quotient questionnaire

(AQ; Baron-Cohen, Wheelwright, Skinner, Martin, & Clubley, 2001);  $t(43.94) = 8.91, p < .001$  and on average, were older than the TD group;  $t(39.54) = 2.09, p = .043$ . For group means and standard deviations on all the measures described above, see Table 1.

(Insert Table 1 here)

## Materials

Sentences from Joseph et al. (2008) were used for this experiment. There were 36 experimental sentences in total, each of which had three versions (for the full stimulus set see Joseph et al., 2008). For an example of the stimuli and for an example of how the sentences were divided into regions of interest for analysis, see Table 2. Two minor adjustments were made to two of the sentences; *evening* was included as a final word in one sentence group and *afternoon* was included as the final word in another sentence group. This was done to create a final region of interest in both sentences, consistent with all the other stimuli. No alteration to the plausibility of the sentences occurred because these minimal changes were at the end of two of the sentences. In the implausible and anomalous sentences, the plausibility violation occurred at the target noun (*milk* in Table 2) that followed the infinitive verb (*to pour/grow* in Table 2). Sentences were matched across conditions such that there were no significant differences in the frequency of the noun prior to the infinitive verb (in Table 2 *bucket/jug/seed*), nor in the frequency and length of the infinitive verb across conditions, and all words following the infinitive verb were exactly the same (Joseph et al., 2008). Three lists of 86 sentences were created, with each list containing a different version of each of the 36 sentences, 40 additional filler sentences and 10 practice sentences that were displayed prior to the experimental stimuli. Each participant only read one of the three lists of sentences.

(Insert table 2 here)

## Design

A 2 X 3 mixed design was employed with group (ASD vs. TD) as a between

participants factor and sentence plausibility (plausible vs. implausible vs. anomalous) as a repeated measures factor.

### **Apparatus**

Participant's eye movements were tracked using an Eyelink 1000 eye tracker (SR Research, Ottawa, Canada) operating at a sampling rate of 1000 Hz, as they read sentences presented on a computer monitor (19 inches). Sentences were displayed in black Courier New 14pt font, with a light grey background. The monitor was set at a distance of 70cm from a headrest that was used to minimize participant movement during testing. Viewing was binocular, but eye tracking was monocular. Forty-three participants had their right eye tracked and three had their left eye tracked.

### **Procedure**

Participants were calibrated using a 3-point sequence of dots that covered the width of the screen in place of where each sentence would appear. Once participants had fixated each calibration point, a validation procedure followed to ensure that each fixation was within 0.50° of each point. Calibration was checked prior to each sentence presentation using a procedure whereby participants had to fixate a dot on the left hand side of the screen where the beginning of each sentence was set to appear. Recalibration was performed if the fixation was off centre.

Participants were warned that some of the sentences might appear "strange" but to read normally. Participants read at their own rate and were instructed to press a button on a controller to indicate when they had finished reading each sentence. Participants were also informed that there would be comprehension questions after approximately half of the sentences, and that they would be required to respond to these by pressing a button to indicate either a *Yes* or *No* response to the question. Instructions reminding participants of which button represented *Yes* and *No* were included underneath each comprehension question.

These questions were factual and did not require detection of anomalies or implausibilities.

These questions were included to ensure that participants read for comprehension. Before the experiment began, ten practice sentences were presented to allow participants to become accustomed to the procedure and to clarify any queries before the experimental materials were presented. The entire eye tracking session lasted approximately 25 minutes. **Data**

### **preparation and analyses**

Sentences were divided into five regions (see Table 2). Of these, three regions were of particular interest; the pre-target region that included the determiner and adjective, except for two stimuli where the pre-target region did not include an adjective, only the determiner ‘the’. We did not edit these sentences, as we did not wish to disrupt sentential context that previously had been pre-screened to result in an implausibility or anomaly. The target region included the critical noun where the plausibility violation occurred and the post-target region that included one long or two short words that immediately followed the target.

These are where disruption in the eye movement record was expected to occur as a result of the plausibility violations.

Sentence comprehension was high and did not differ between ASD and TD groups (TD  $M = 0.97$ ,  $SD = 0.03$ ; ASD  $M = 0.96$ ,  $SD = 0.04$ ), with all participants correctly answering at least 86% of questions  $t(40.04) = 0.67$ ,  $p = 0.50$ . A default cleaning process for reading experiments was carried out in DataViewer (SR Research, Ottawa, Canada), whereby contiguous fixations that had a duration of 80ms or less and were within  $.50^\circ$  of one another were merged. Fixations were also merged in instances when there were three or more contiguous fixations, each less than 140ms within a region. Fixations below 80ms are unlikely to result in meaningful information being extracted from the text and fixations above 800ms are likely to be a result of tracker error and were therefore removed, resulting in a data loss of 3.49% (ASD = 1.74%, TD = 1.75%).



Trials were also excluded if there was tracker loss, if a participant blinked whilst fixating the target region, if participants failed to fixate at least two of the three ROI's, or if the trial had been disrupted during the testing session e.g., participant talking to the experimenter. These exclusions resulted in a total loss of 11.54% of experimental trials (ASD = 7.46%, TD = 4.08%).

For each ROI, each of the following eye movement measures were examined: first fixation duration (the duration of the first fixation made in a region), single fixation duration (the duration of a fixation in a region when this is the only fixation made in that region) and gaze duration (the sum of all fixations in a region from the first fixation on the region until the eyes leave the region from either left or right). These measures are usually taken to reflect early stages of linguistic processing in reading. We also analysed go past time (the sum of all fixations from the first fixation in a region until the eyes leave the region to the right, including any regressive fixations made to prior areas of the text) and total time (the sum of all fixations in a region), both of which are taken to reflect somewhat later stages of processing.

Data points from each eye movement measure were removed if more than 2.5 standard deviations away from the group by condition mean, which resulted in a loss of no more than 3.71% of data from each measure (approximately equal proportions of data were removed across groups for each measure). Each of the eye movement measures were log transformed and linear mixed effect models (Baayen, Davidson, & Bates, 2008) were computed in R (version 3.2.4; R Core Team, 2016) using the lme4 package (Bates, Maechler, Bolker, & Walker, 2015). The model computed for each measure examined whether there was a difference between groups, sentence plausibility or any interactions. Group and plausibility were both specified as categorical fixed effects and deviation contrasts were coded to examine whether there was a difference between ASD and TD readers, using the

contr.sdif function from the MASS library (Venables & Ripley, 2002). In addition, two deviation contrasts were manually coded to examine the difference between anomalous and plausible sentences (anomalous -.5, implausible, 0, plausible, .5) and implausible and plausible sentences (anomalous 0, implausible -.5, plausible, .5). This user specified contrast matrix was inversed for analysis using the ginv function from the MASS library (Venables & Ripley, 2002). Age was included (centered) as a continuous fixed effect, in order to control for the age difference between ASD and TD readers and assure that any effects of group were a result of ASD. As is recommended, the full random structure was included (Barr, Levy, Scheepers & Tily, 2013), which meant that crossed random effects were included for participants and items, with random slopes for sentence plausibility at the participant level and random slopes for sentence plausibility, group and centered age at the item level. This resulted in the following syntax; `Model = lmer(logDV ~ group* plausibility + centered_age + (1 + plausibility | participant_id) + (1 + group* plausibility + centered_age | item_id), data = data)`. The lmerTest package was used to compute *p* values (Kuznetsova, Brockhoff & Christensen, 2016). If a model would not converge, parameters were incrementally removed from the random structure, beginning with the items level. The model was initially re-run excluding the correlation. If this was unsuccessful, the correlation was re-entered and the model was re-run excluding the interaction. If the model would still not converge, a model excluding both the correlation and interaction was run, before removing random slopes one by one (age followed by condition followed by group). If the model would still not converge and only the random intercept for items remained, the correlation at the participant level was then removed, followed by the random slope. Prior to examining and interpreting model output, the distribution and normality of residuals (Pinheiro & Bates, 2000) was examined using QQ and density plots. The output for all fixed effect parameters can be viewed in the online appendix, in addition to raw data and R code used to compute analyses.

## Results

### Global measures

Before examining the effect of plausibility for the regions of interest, we examined whether there were any basic sampling differences between the ASD and TD groups. To do this, three global processing measures were analysed; mean fixation duration (the mean fixation duration calculated from all fixations in a trial), number of fixations (the sum of fixations made during a trial), and sentence reading time (time from trial onset until participants made a manual response). Means and standard deviations of these measures are included in Table 3. For clarity and succinctness, interactions are only reported if reliable. In addition, age did not have a reliable effect on any of the measures reported below, and therefore will not be discussed.

Analysis of mean fixation duration data indicated that there was no effect of group  $b = -0.04$ ,  $SE = 0.04$ ,  $t = 1.14$ ,  $p = .259$  and no difference between mean fixation durations for implausible  $b < 0.01$ ,  $SE = < 0.01$ ,  $t = 0.35$   $p = .726$  or anomalous sentences  $b = < 0.01$ ,  $SE = < 0.01$ ,  $t = 0.93$ ,  $p = .355$ , in comparison to the plausible sentences. For fixation count, there was a numerical trend to suggest ASD readers made more fixations than the TD group, but this was not reliable  $b = -0.20$ ,  $SE = 0.10$ ,  $t = 1.92$   $p = .061$ . However, there was an effect of sentence plausibility, with both TD and ASD readers making more fixations when reading anomalous sentences  $b = -0.08$ ,  $SE = 0.01$ ,  $t = 6.89$ ,  $p < .001$  in comparison to plausible sentences, but there was no difference between plausible and implausible sentences  $b = -0.01$ ,  $SE = 0.01$ ,  $t = 1.47$ ,  $p = .151$ . Consistent with the numerical group effect for fixation count, analyses for sentence reading times indicated that ASD readers had longer sentence reading times overall  $b = -0.26$ ,  $SE = 0.12$ ,  $t = 2.22$ ,  $p = .032$ , and both groups had longer reading times for implausible  $b = -0.02$ ,  $SE = 0.01$ ,  $t = 2.18$ ,  $p = .036$  and anomalous sentences  $b = -0.08$ ,  $SE = 0.01$ ,  $t = 7.26$ ,  $p < .001$  in comparison to plausible sentences.

From these global measures, we can tentatively conclude that there were no overall differences in the speed with which ASD and TD groups' extracted information from the text within individual fixations. The ASD readers did however have longer reading times overall, in comparison to the TD group. This is consistent with our previous studies examining reading in ASD that have found increased rereading behaviour for these individuals (Au-Yeung, Kaakinen, Liversedge & Benson, 2015; Howard et al., 2016). Importantly, there were no reliable interactions, with both groups showing comparable global effects of anomaly and implausibility upon the number of fixations made and sentence reading time. This indicates that these manipulations had a comparable overall impact upon language processing for both TD and ASD readers. Next, we will consider the fine-grained measures to examine the time course of anomaly and implausibility detection and processing in both groups.

(Insert Table 3 here)

### **Pre-target region**

At the pre-target region all sentence types were plausible, and as such, no differences between groups or sentence types were expected in early processing measures. For means and standard deviations for all pre-target, target and post target measures, see Table 4. Consistent with our expectations, no group differences were reliable for first fixation duration, single fixation duration, gaze durations or go past time ( $ts < 0.58$ ,  $ps > .567$ ). In addition, sentence plausibility also had no reliable effect on the duration of first fixations, single fixations, or gaze durations ( $ts < 1.64$ ,  $ps > .102$ ). However, sentence plausibility did affect the duration of go past times, with both groups taking significantly longer to proceed past the pre-target region when the upcoming target word was anomalous, in comparison to plausible  $b = -0.06$ ,  $SE = 0.01$ ,  $t = 4.23$ ,  $p < .001$ , but no difference was found between the implausible and plausible sentences  $b = -0.02$ ,  $SE = 0.01$ ,  $t = 1.74$ ,  $p = .087$ .

The effect of anomaly on go past times was not predicted because at this point in

time, participants had not yet fixated the target region where the plausibility violation occurred. However, this effect has been previously reported for experiments that have manipulated the plausibility of target words (e.g., Rayner et al., 2004) and there are two possible explanations. Firstly, it could be argued that the increased go past time for anomalous sentences are a parafoveal-on-foveal-effect. Such effects occur when the semantic characteristics of an upcoming word ( $n+1$ ) influence the processing of the currently fixated word or region ( $n$ ). This explanation is consistent with models of eye movement control during reading whereby attention is graded and permits the identification of multiple words in parallel (e.g., SWIFT; Engbert, Nuthmann, Richter & Kliegl, 2001). Alternatively, this effect might occur as a result of saccadic undershoots or small calibration errors that result in attention being allocated to the target word, but fixations located (or detected to be located) on the pre-target word (for a detailed discussion the mislocated fixations account, see Drieghe, Rayner & Pollatsek, 2008). This explanation is consistent with models of reading that predict attention to be allocated serially, with only low level information such as orthography and phonology being extracted from words in the parafovea (e.g., E-Z reader; Reichle, Rayner & Pollatsek, 2003). The exact cause of this effect, and whether it is evidence of parallel processing or something more trivial with regard to oculomotor or tracker error is not of critical concern for this experiment. What is important for this experiment is that this effect was constant across our groups, indicating that the processing of the pre-target region was comparable for TD and ASD readers.

### **Target region**

The target region was the word at which the plausibility violation occurred and disruption in the eye movement record was expected. Both first fixation and single fixation duration data showed the same pattern of results. No differences were found between the groups for first  $b = -0.04$ ,  $SE = 0.05$ ,  $t = 0.83$ ,  $p = .410$  or single fixation duration  $b = -0.07$ ,

$SE = 0.05, t = 1.22, p = .229$ , but there was a significant increase in both first  $b = -0.04, SE = 0.01, t = 3.94, p < .001$  and single fixation durations  $b = -0.06, SE = 0.01, t = 4.18, p < .001$  upon the target when it was anomalous, in comparison to plausible. This suggests that during the earliest stages of foveal processing of the target, both TD and ASD participants detected the anomalies. No overall effect of implausibility was detected for first fixation durations  $b = -0.01, SE = 0.01, t = 1.39, p = .172$ , but the effect of implausibility was reliable for single fixation durations  $b = -0.04, SE = 0.01, t = 2.69, p = .009$ . However, this was qualified by a significant interaction between group and implausibility for both first  $b = -0.05, SE = 0.02, t = 2.41, p = .017$  and single fixation durations  $b = -0.06, SE = 0.02, t = 2.32, p = .021$ .

In order to examine the nature of this interaction, the model was re-run separately for each group and each measure. The results indicated that for both first and single fixation durations, TD readers had longer fixation durations for both anomalous  $b = -0.05, SE = 0.01, t = 3.66, p < .001$ ;  $b = -0.07, SE = 0.02, t = 3.82, p = .001$  and implausible target words  $b = -0.04, SE = 0.01, t = 2.91, p = .006$ ;  $b = -0.06, SE = 0.02, t = 3.82, p = .001$ , in comparison to plausible target words. In contrast, the ASD readers showed an increase in first and single fixation time for anomalous target words in comparison to the plausible  $b = -0.03, SE = 0.02, t = 1.84, p = .076$ ;  $b = -0.05, SE = 0.02, t = 2.04, p = .050$  (marginal for first fixation), but no difference between implausible and plausible targets  $b = 0.01, SE = 0.02, t = 0.36, p = .720$ ;  $b = -0.01, SE = 0.02, t = 0.36, p = .722$ . These analyses indicate that both TD and ASD participants detected the anomalies upon initial fixation. However, detection of the implausibilities was present in the TD group, but absent for the ASD readers (See Figure 1).

(Insert Figure 1 here)

There was no difference in gaze duration between groups  $b = -0.01, SE = 0.05, t = 0.26, p = .799$ , but a reliable effect of sentence plausibility was detected, with again both groups having longer gaze durations upon anomalous target words, in comparison to

plausible target words  $b = -0.06$ ,  $SE = 0.01$ ,  $t = 4.44$   $p < .001$ , but no overall difference between implausible and plausible target words  $b = -0.03$ ,  $SE = 0.02$ ,  $t = 1.69$ ,  $p = .102$ . The interaction term between the effect of implausibility and group was not significant for this measure  $b = -0.04$ ,  $SE = 0.03$ ,  $t = 1.59$ ,  $p = .119$ . There was no difference overall between go past times for the TD and ASD groups  $b = -0.05$ ,  $SE = 0.07$ ,  $t = 0.81$ ,  $p = .425$ . Both groups had longer go past times when the target word was anomalous  $b = -0.10$ ,  $SE = 0.02$ ,  $t = 5.24$ ,  $p < .001$  and there was a marginal effect of implausibility  $b = -0.03$ ,  $SE = 0.02$ ,  $t = 1.98$ ,  $p = .054$ .

To summarise the findings for the target region; there were no reliable differences in the speed with which the TD and ASD groups detected the anomalies. Both groups detected anomalies very rapidly, as indexed by increased first and single fixation durations upon anomalous target words, relative to the plausible. The disruption to reading as a result of the implausibilities in these early measures, however, was only evident for the TD group. The TD group detected implausibilities as rapidly as anomalies, with first and single fixation durations being inflated. In contrast, the ASD group did not show disruption for any of the early stages of target word processing, as a result of the implausibility manipulation. This finding suggests that the ASD group did not detect the implausibilities during initial target word processing.

### **Post target region**

The post target region included the words that immediately followed the target region. For first fixation duration, single fixation duration, gaze duration and go past time there was no reliable difference between the TD and ASD groups ( $ts < 1.21$ ,  $ps > .233$ ). There was also no reliable effect of sentence plausibility for first fixation durations, single fixation durations or gaze durations ( $ts < 1.48$ ,  $ps > .148$ ). However, an effect of anomaly was present for go past time  $b = -0.11$ ,  $SE = 0.03$ ,  $t = 4.19$ ,  $p < .001$ , which indicates that participants spent

longer rereading previous areas of the anomalous sentences in comparison to the plausible, prior to making a rightward saccade out of the post target region. There was no difference between go past times for implausible and plausible sentences  $b = -0.02$ ,  $SE = 0.02$ ,  $t = 0.96$ ,  $p = .340$ . Together these results suggest that the disruption experienced when readers initially encountered the anomalies and implausibilities in the target region did not spill over and affect early processing of the words that followed, but the anomalies did result in increased go past times.

### **Total times**

Total time includes all fixations in a region, including those made during second pass reading (the period of time after the text has been read through once in entirety). For the pre-target region, ASD readers had increased total times, in comparison to TD readers  $b = -0.24$   $SE = 0.11$ ,  $t = 2.13$ ,  $p = .039$ . In addition, the total time spent in the pre-target region was affected by sentence plausibility, with longer total times occurring for both groups when the sentence was anomalous in comparison to plausible  $b = -0.19$   $SE = 0.02$ ,  $t = 9.25$ ,  $p < .001$ , but the implausible and plausible sentences did not differ from one another  $b = -0.02$   $SE = 0.02$ ,  $t = 0.93$ ,  $p = .361$ .

In the target region there was a difference between the TD and ASD groups total times  $b = -0.20$   $SE = 0.09$ ,  $t = 2.15$ ,  $p = .038$ , with the ASD group spending longer fixating this region overall. There was also a reliable effect of plausibility in the target region, with participants spending significantly longer in this region when the sentence was anomalous in comparison to plausible  $b = -0.13$   $SE = 0.02$ ,  $t = 6.07$ ,  $p < .001$ , but the implausible and plausible sentences did not differ  $b = -0.02$   $SE = 0.02$ ,  $t = 1.14$ ,  $p = .263$ .

In the post-target region, there was no reliable effect of group  $b = -0.17$   $SE = 0.11$ ,  $t = 1.54$ ,  $p = .130$ , but there was a reliable effect of sentence type, and identical to the findings for the previous regions, participants spent significantly longer amounts of time fixating the



post target region when the sentences were anomalous in comparison to plausible  $b = -0.08$   $SE = 0.02$ ,  $t = 3.80$ ,  $p = .001$ , but the implausible and plausible sentences did not differ  $b = -0.02$   $SE = 0.02$ ,  $t = 0.98$ ,  $p = .329$ .

(Insert Table 4 here)

### Supplementary Analysis

Considering the lack of difference between the ASD and TD groups first pass reading times (gaze durations), yet clear differences in total and sentence reading times, it seemed reasonable to explore the nature of this increased reading time in ASD. In the following supplementary analyses, we were keen to establish the time course of this increased rereading in the ASD group and whether this was localised to a particular ROI. If the increased total times for the ASD readers arose due to a higher proportion of regressive fixations during first pass reading of the sentences, then this might suggest that they experienced difficulty constructing an initial interpretation of the sentence. Alternatively, if rereading occurred during second pass (or later) reading, then this might indicate that whilst ASD readers did not differ from the TD group in their initial construction of an interpretation of the sentence, their evaluation of this interpretation caused them to reread the sentences. Means and standard deviations for all supplementary analyses are presented in Table 5. Below we only report group differences and interactions (if reliable). Those interested in how sentence type mediated these differences across groups are referred to the online supplementary material where full model output is presented.

Firstly, we examined the proportion of first pass regressions made out of each ROI (prior to a reader fixating information to the right of a ROI). This was to identify the time course of rereading, in other words, whether the increased rereading for ASD participants occurred during first pass of the sentence (prior to a participant proceeding to fixate new rightward information). No differences between the proportions of first pass regressions

made out of the pre-target, target or post target regions were found ( $z$ s  $< 1.10$ ,  $p$ s  $> .274$ ).

Thus, ASD readers were no more likely than TD readers to regress in order to reread during the first pass through the sentence.

Secondly, we examined the proportion and duration of rereading (total time minus gaze duration), in order to examine whether a particular region rereading for ASD participants was localised. For the pre-target region the ASD group were found to engage in rereading on a higher proportion of trials in comparison to TD readers  $b = -0.80$   $SE = 0.37$ ,  $z = 2.16$ ,  $p = .031$ , but no difference was found between the groups for duration of rereading  $b = -0.23$   $SE = 0.13$ ,  $t = 1.76$ ,  $p = .086$ . Similarly, for the target region ASD readers were found to reread on a higher proportion of trials, in comparison to TD readers  $b = -0.88$   $SE = 0.37$ ,  $z = 2.39$ ,  $p = .017$ . However, there was no evidence that there was any difference in the amount of time the two groups spent rereading when they revisited this region  $b = -0.08$   $SE = 0.10$ ,  $t = 0.82$ ,  $p = .419$ . For the post target region, a marginal difference between groups was found for the proportion of rereading  $b = -0.63$   $SE = 0.35$ ,  $z = 1.82$ ,  $p = .069$ , with a trend suggesting ASD participants revisited this region to reread on a higher proportion of trials than TD participants. There was also a marginal difference between groups for rereading duration  $b = -0.21$   $SE = 0.11$ ,  $t = -1.94$ ,  $p = .060$ , indicating that there was a tendency for ASD readers to spend longer rereading information in the post-target region too.

### Discussion

The on-line use of world knowledge during reading in ASD was examined by monitoring the eye movements of participants as they read sentences that were plausible, implausible or anomalous. Both the TD and ASD groups detected the anomalies almost immediately, as indexed by increased first fixation durations on the target word. The anomalies also disrupted later sentence processing in both groups, as indexed by go past times for the target and post target region, and total times for all critical regions. The TD

group detected the implausibilities as rapidly as the anomalies, with disruption occurring during first fixations on the target word and this effect was also evident for single fixation durations. This is the first study to report that TD readers detect implausibilities as rapidly as anomalies. Previous studies have reported disruption as a result of implausibilities in later measures (Rayner et al., 2004; Joseph et al., 2008). However, the disruption as a result of the implausibilities in the current study was shorter lived in comparison to the disruption as a result of the anomalies, and was only evident in these very early measures. Therefore, although the speed of detection is inconsistent with previous findings, the reduced disruption as a result of implausibilities relative to anomalies is comparable (e.g., Joseph et al. 2008; Rayner et al. 2004). One possible explanation for the differential time course of implausibility effects for the TD readers in this and previous studies is the age of our TD participants. Previous studies that have used similar manipulations have recruited undergraduate students who are approximately 18 years of age, and Joseph et al. (2008) used a similar sample when norming the experimental stimuli that we adopted here. In the current study individuals were recruited from the local community and had an average age of 29 years. It's therefore possible that the increased life and language experience of our participants resulted in them detecting implausibilities more immediately than has been previously reported for undergraduate readers.

Our critical finding was the interaction between group and the effect of implausibility for first and single fixation durations in the target region. This revealed that the ASD readers, unlike TD readers, failed to detect implausibilities upon initial fixation of the target. Moreover, the ASD readers did not appear to show any disruption as a result of the implausibilities at any point during the processing of the critical regions. Disruption was found however for the global measure of sentence reading time, which indicates that ASD readers did detect and experience disruption to reading as a result of the implausible semantic

oddities relative to the plausible sentences. The time course difference indicates that the detection of implausibilities was delayed for ASD readers, relative to TD readers.

This finding partially supports our predictions. Based on the assumption that both types of linguistic manipulation require the evaluation of world knowledge for the oddities to be detected, we predicted that there would be a delay in the detection of both anomalies and implausibilities in ASD. However, we found ASD readers to be delayed in the detection of implausibilities, but not in anomaly detection. Recall that the anomalous sentences not only violated world knowledge, but also violated a verb's selectional restrictions, which are semantic rules about what can and cannot be an argument to the verb. This information is activated when a verb is lexically identified and is then used to assign thematic roles (e.g., Carlson & Tanenhaus, 1988). Therefore, it is possible that the reason the participants with ASD immediately detected the anomalies, but not the implausibilities, is because the anomalies could be detected without the use of world knowledge, on the basis of information activated during relatively early lexical stages of processing (e.g., Warren & McConnell, 2007). In contrast, the evaluation of world knowledge was critical for the detection of implausibilities that were not detectable based on verb argument violations. What this means in relation to our hypothesis is that the detection of semantic oddities that require the evaluation of world knowledge is less efficient (delayed) in ASD, but the detection of semantic anomalies which are a result of verb-argument violations, and which may be detected on the basis of selectional restriction information, is not.

One might consider these results to be in conflict with Howard et al.'s (2016) finding that ASD readers detected that they had misinterpreted an ambiguous prepositional phrase, as quickly as TD readers, based upon their evaluation using world knowledge. However, a closer look at the materials used in the Howard et al. (2016) experiment indicates that a high proportion (70%) of the stimuli were anomalous as a result of a violation of a verb's

selectional restrictions. Therefore, it is possible that the quite immediate disruption seen for ASD readers by Howard et al. (2016) was not evidence for intact world knowledge use, but instead evidence in support of intact detection of selectional restriction violations at an earlier stage of processing.

What our data very clearly demonstrate is that consistent with what CIP theory predicts, when the use of world knowledge is required during reading, subtle differences in the time-course of sentence processing are apparent for TD and ASD readers. These findings are in contrast with Saldaña and Frith's (2009) conclusion that the speed and access to world knowledge during reading in ASD is as efficient as TD readers. Our results are also inconsistent with Sansosti et al.'s (2013) finding that readers with ASD have longer fixation durations. It is possible that the difference between our own and Sansosti et al.'s (2013) findings may be related to differences in the stimuli they used, which required an inference to be computed, a demand that was not required in our own study and one which may have induced such processing differences. We did however replicate the finding that overall readers with ASD take longer to read sentences (Au-Yeung et al., 2015; Howard et al., 2017; Sansosti et al. 2013).

The supplementary analyses demonstrated that ASD readers revisit each ROI to reread on a higher proportion of trials than TD readers. Moreover, this rereading did not appear to be localised to any particular ROI but reflected the ASD participants rereading once the sentences had been read through entirely. The lack of difference found in first pass regressions and lack of group interactions specific to the rereading of anomalous and implausible sentences suggests that this increased rereading is not a result of a linguistic processing difference *per se*. Thus, the rereading ASD participants engaged in may have been related to the evaluation of their initial interpretation. This idea is consistent with what has

been previously reported, when the time-course of rereading in ASD has been examined (Howard et al., 2017).

It is possible that the inclusion of comprehension questions may have led the ASD group to be especially aware of the requirement to comprehend the sentences correctly, leading these readers to be more hesitant to press a button and confirm that they had finished reading each sentence. Note that we are not arguing that the ASD participants are simply slower to react. Instead, we are suggesting that it may take ASD participants longer to develop a sense of confidence in relation to any response they may make about their interpretation of what they have just read. The sensitivity of ASD groups to instruction requirements and task demands is increasingly recognised in the literature to be a factor that affects performance on tasks assessing aspects of cognitive processing (e.g., see the review of performance on executive functioning tasks in White, 2013). It is also noteworthy that several of the participants with ASD who took part in this experiment vocalized anxieties about the prospect of answering comprehension questions, indicating that this was a task they had had difficulty with in the past. Therefore, the possibility that the increased rereading in our ASD sample reflects an increased ‘checking’ of an interpretation of a sentence as a result of apprehension concerning upcoming comprehension questions, seems potentially reasonable, but remains to be empirically tested. Similar reports of repeated sampling of task relevant information has also been recently observed during scene inspection in ASD (Benson, Castelhana, Au-Yenug & Rayner, 2012; Benson, Castelhana, Howard, Latif & Rayner, 2015).

We have championed the use of eye tracking to examine language processing in this paper, because of its capacity to provide detailed information about on-line language processing in ASD. We realise that this is an indirect measure of world knowledge processing, but we believe that this method clearly provides much more information

processing detail in comparison to traditional RT and Accuracy measures. Further research using methods that examine both on-line behavioural measures and neural activity, for example, examining fixation related potentials through the co-registration of eye tracking and EEG would provide insight into the qualitative differences in the neural systems that underlie temporal processing differences in language processing in ASD. Since the current impact of this research is predominantly theoretical, this could be noted as a limitation to the work, however, these findings and the research that they subsequently motivate, have potential to contribute to the development of more effective application techniques and guidelines in relation to reading development and comprehension in ASD.

To conclude, differences in the speed with which world knowledge was used in written language processing were present between an ASD and TD group when reading single sentences containing implausibilities. ASD readers did, however, detect anomalies that were a result of selectional restriction violations as quickly as TD readers. Thus, this study demonstrates both that there are subtle differences in the time course with which world knowledge is used to evaluate sentence meaning during reading in ASD. It would seem reasonable to conclude that the performance differences found in ASD groups during higher order linguistic tasks may in part be a consequence of less efficient world knowledge processing.

## Ethical Statement

The University of Southampton's Ethics and Research Governance Board ethically approved this research and all participants gave informed written consent before taking part.

The authors declare that they have no conflict of interest.



## References

- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Arlington, VA: American Psychiatric Publishing.
- Åsberg, J., & Sandberg, A. D. (2012). Dyslexic, delayed, precocious or just normal? Word reading skills of children with autism spectrum disorders. *Journal of Research in Reading*, 35(1), 20-31.
- Åsberg, J., Kopp, S., Berg-Kelly, K., & Gillberg, C. (2010). Reading comprehension, word decoding and spelling in girls with autism spectrum disorders (ASD) or attention-deficit/hyperactivity disorder (AD/HD): performance and predictors. *International Journal of Language & Communication Disorders*, 45(1), 61-71.
- Au-Yeung, S. K., Kaakinen, J. K., Liversedge, S. P., & Benson, V. (2015). Processing of written irony in autism spectrum disorder: An eye movement study. *Autism Research*, 8, 749–760. <http://dx.doi.org/10.1002/aur.1490>
- Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of memory and language*, 59(4), 390-412.
- Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001). The Autism-Spectrum Quotient (AQ): Evidence from Asperger syndrome/high-functioning autism, males and females, scientists and mathematicians. *Journal of Autism and Developmental Disorders*, 31(1), 5-17.
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68, 255–278. <http://dx.doi.org/10.1016/j.jml.2012.11.001>
- Bates, D., Maechler, M., Bolker, B., Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1), 1-48. doi:10.18637/jss.v067.i01.

- Benson, V., Castelhana, M. S., Au-Yeung, S. K., & Rayner, K. (2012). Eye movements reveal no immediate "WOW" ("which one's weird") effect in autism spectrum disorder. *Quarterly Journal of Experimental Psychology*, 65(6), 1139-1150.
- Benson, V., Castelhana, M. S., Howard, P. L., Latif, N., & Rayner, K. (2015). Looking, seeing and believing in autism: Eye movements reveal how subtle cognitive processing difference impact in the social domain. *Autism Research*, doi: 10.1002/aur.1580
- Bodner, K. E., Engelhardt, C. R., Minshew, N. J., & Williams, D. L. (2015). Making inferences: Comprehension of physical causality, intentionality, and emotions in discourse by high-functioning older children, adolescents, and adults with autism. *Journal of Autism and Developmental Disorders*, 45(9), 2721-2733.
- Brock, J., & Bzishvili, S. (2013). Deconstructing Frith and Snowling's homograph-reading task: Implications for autism spectrum disorders. *Quarterly Journal of Experimental Psychology*, 66(9), 1764-1773.
- Brock, J., & Caruana, N. (2014). Reading for sound and reading for meaning in autism: Frith and Snowling (1983) revisited. In J. Arciuli & J Brock (Eds.), *Communication in autism* (pp. 125–145). Amsterdam, the Netherlands: John Benjamins.
- Brock, J., Norbury, C., Einav, S., & Nation, K. (2008). Do individuals with autism process words in context? Evidence from language-mediated eye-movements. *Cognition*, 108(3), 896-904.
- Brown, H. M., Oram-Cardy, J., & Johnson, A. (2013). A Meta-Analysis of the Reading Comprehension Skills of Individuals on the Autism Spectrum. *Journal of Autism and Developmental Disorders*, 43(4), 932-955.
- Carlson, G.N., & Tanenhaus, M. K. (1988). Thematic roles and language comprehension. In W. Wilkins, (Ed), *Thematic Relations* (pp. 263-288). New York: Academic Press.

- Caruana, N., & Brock, B. (2014). No association between autistic traits and contextual influences on eye-movements during reading. *PeerJ*, 2:e446.
- Conti-Ramsden, G., Botting, N., & Faragher, B. (2001). Psycholinguistic markers for specific language impairment (SLI). *Journal of Child Psychology and Psychiatry*, 42(6), 741-748.
- Dennis, M., Lazenby, A. L., & Lockyer, L. (2001). Inferential language in high-function children with autism. *Journal of Autism and Developmental Disorders*, 31(1), 47-54.
- Drieghe, D., Rayner, K., & Pollatsek, A. (2008). Mislocated fixations can account for parafoveal-on-foveal effects in eye movements during reading. *The Quarterly Journal of Experimental Psychology*, 61(8), 1239-1249.
- Engbert, R., Nuthmann, A., Richter, E. M., & Kliegl, R. (2005). SWIFT: a dynamical model of saccade generation during reading. *Psychological review*, 112(4), 777.
- Ferguson, H. J., & Sanford, A. J. (2008). Anomalies in real and counterfactual worlds: An eye-movement investigation. *Journal of Memory and Language*, 58(3), 609-626.
- Frith, U. (1989). *Autism: Explaining the enigma* (1989). Blackwell Scientific Publications: Oxford.
- Frith, U. (2012). The 38th Sir Frederick Bartlett Lecture. Why we need cognitive explanations of autism. *Quarterly Journal of Experimental Psychology*, 65(11), 2073-2092.
- Frith, U., & Happé, F. (1994). Autism - Beyond the theory of mind. *Cognition*, 50(1-3), 115-132.
- Frith, U., & Snowling, M. (1983). Reading for meaning and reading for sound in autistic and dyslexic children. *British Journal of Developmental Psychology*, 1, 329-342.

- Gernsbacher, M. A. (1991). Cognitive-processes and mechanisms in language comprehension - The structure building framework. *Psychology of Learning and Motivation-Advances in Research and Theory*, 27, 217-263.
- Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review*, 101(3), 371-395.
- Hahn, N., Snedeker, J., & Rabagliati, H. (2015). Rapid linguistic ambiguity resolution in young children with autism spectrum disorder: Eye tracking evidence for the limits of weak central coherence. *Autism Research*, 8, 717–726.
- Hala, S., Pexman, P. M., & Glenwright, M. (2007). Priming the meaning of homographs in typically developing children and children with autism. *Journal of Autism and Developmental Disorders*, 37(2), 329-340.
- Happé, F. G. E. (1997). Central coherence and theory of mind in autism: Reading homographs in context. *British Journal of Developmental Psychology*, 15, 1-12.
- Happé, F., & Frith, U. (2006). The weak coherence account: Detail-focused cognitive style in autism spectrum disorders. *Journal of Autism and Developmental Disorders*, 36(1), 5-25.
- Henderson, L. M., Clarke, P. J., & Snowling, M. J. (2011). Accessing and selecting word meaning in autism spectrum disorder. *Journal of Child Psychology and Psychiatry*, 52(9), 964-973.
- Howard, P. L., Liversedge, S. P., & Benson, V. (2016). Benchmark Eye Movement Effects During Natural Reading in Autism Spectrum Disorder. *Journal of Experimental Psychology: Learning, Memory & Cognition*.
- Huemer, S. V., & Mann, V. (2010). A Comprehensive Profile of Decoding and Comprehension in Autism Spectrum Disorders. *Journal of Autism and Developmental Disorders*, 40(4), 485–493.

- Jolliffe, T., & Baron-Cohen, S. (1999). A test of central coherence theory: linguistic processing in high-functioning adults with autism or Asperger syndrome: is local coherence impaired? *Cognition*, 71(2), 149-185.
- Jolliffe, T., & Baron-Cohen, S. (2000). Linguistic processing in high-functioning adults with autism or Asperger's syndrome. Is global coherence impaired? *Psychological Medicine*, 30(5), 1169-1187.
- Jones, C. R. G., Happe, F., Golden, H., Marsden, A. J. S., Tregay, J., Simonoff, E., . . . Charman, T. (2009). Reading and Arithmetic in Adolescents With Autism Spectrum Disorders: Peaks and Dips in Attainment. *Neuropsychology*, 23(6), 718-728.
- Joseph, H. S. S. L., Liversedge, S. P., Blythe, H. I., White, S. J., Gathercole, S. E., & Rayner, K. (2008). Children's and adults' processing of anomaly and implausibility during reading: Evidence from eye movements. *Quarterly Journal of Experimental Psychology*, 61(5), 708-723.
- Just, M. A., Cherkassky, V. L., Keller, T. A., & Minshew, N. J. (2004). Cortical activation and synchronization during sentence comprehension in high-functioning autism: evidence of underconnectivity. *Brain*, 127, 1811-1821.
- Just, M. A., Cherkassky, V. L., Keller, T. A., Kana, R. K., & Minshew, N. J. (2007). Functional and anatomical cortical underconnectivity in autism: Evidence from an fMRI study of an executive function task and corpus callosum morphometry. *Cerebral Cortex*, 17(4), 951-961.
- Kintsch, W. (1988). The role of knowledge in discourse comprehension - A construction integration model. *Psychological Review*, 95(2), 163-182.
- Kuznetsova, A., Brockhoff, A. B., & Christensen, R. H. B. (2016). lmerTest: Tests in Linear Mixed Effects Models. R package version 2.0-30. <https://CRAN.R-project.org/package=lmerTest>

- Liversedge, S. P., & Findlay, J. M. (2000). Saccadic eye movements and cognition. *Trends in Cognitive Sciences*, 4(1), 6-14.
- López, B., & Leekam, S. R. (2003). Do children with autism fail to process information in context? *Journal of Child Psychology and Psychiatry*, 44(2), 285-300.
- Lord, C., Rutter, M., DiLavore, P. C., Risi, S. (2001). Autism Diagnostic Observation Schedule. Los Angeles: Western Psychological Services; 2001.
- Lord, C., Rutter, M., & Le Couteur, A. (1994). Autism Diagnostic Interview-Revised: a revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *Journal of autism and developmental disorders*, 24(5), 659-685.
- Lucas, R., & Norbury, C. F. (2014). Levels of text comprehension in children with Autism Spectrum Disorders (ASD): the influence of language phenotype. *Journal of Autism and Developmental Disorders*, 44(11), 2756-2768.
- Lucas, R., & Norbury, C. F. (2015). Making inferences from text: it's vocabulary that matters. *Journal of Speech, Language, and Hearing Research*, 58(4), 1224-1232.
- Mayes, S. D., & Calhoun, S. L. (2006). Frequency of reading, math, and writing disabilities in children with clinical disorders. *Learning and Individual Differences*, 16(2), 145-157.
- McKoon, G., & Ratcliff, R. (1992). Inference during reading. *Psychological Review*, 99(3), 440-466.
- Minshew, N. J., & Goldstein, G. (1998). Autism as a disorder of complex information processing. *Mental Retardation and Developmental Disabilities Research Reviews*, 4(2), 129-136.
- Minshew, N. J., Goldstein, G., & Siegel, D. J. (1995). Speech and language in high-functioning autistic individuals. *Neuropsychology*, 9(2), 255-261.

- Minschew, N. J., Goldstein, G., & Siegel, D. J. (1997). Neuropsychologic functioning in autism: Profile of complex information processing disorder. *Journal of the International Neuropsychological Society*, 3(4), 303-316.
- Minschew, N. J., Williams, D. L., & McFadden, K. (2008). Information Processing, Neural Connectivity, and Neuronal Organization. In A. W. Zimmerman (Ed.), *Autism: Current Theories and Evidence* (pp. 381-405).
- Nation, K., Clarke, P., Wright, B., & Williams, C. (2006). Patterns of reading ability in children with autism spectrum disorder. *Journal of Autism and Developmental Disorders*, 36(7), 911-919.
- Newman, T. M., Macomber, D., Naples, A. J., Babitz, T., Volkmar, F., & Grigorenko, E. L. (2007). Hyperlexia in children with Autism Spectrum Disorders. *Journal of Autism and Developmental Disorders*, 37, 760-774.
- Norbury, C. F. (2005). Barking up the wrong tree? Lexical ambiguity resolution in children with language impairments and autistic spectrum disorders. *Journal of Experimental Child Psychology*, 90, 142-171.
- Norbury, C. F., & Bishop, D. V. M. (2002). Inferential processing and story recall in children with communication problems: a comparison of specific language impairment, pragmatic language impairment and high-functioning autism. *International Journal of Language & Communication Disorders*, 37(3), 227-251.
- Norbury, C., & Nation, K. (2011). Understanding Variability in Reading Comprehension in Adolescents With Autism Spectrum Disorders: Interactions With Language Status and Decoding Skill. *Scientific Studies of Reading*, 15(3), 191-210.
- Pinheiro, J. C., & Bates, D. M. (2000). *Mixed-Effects Models in S and S-PLUS*. Springer; New York.

- R Core Team. (2016). R: language and environment for statistical computing. R Foundation for Statistical Computing, 2005; Vienna, Austria.
- Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 124(3), 372-422.
- Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *Quarterly Journal of Experimental Psychology*, 62(8), 1457-1506.
- Rayner, K., Carlson, M., & Frazier, L. (1983). The interaction of syntax and semantics during sentence processing: Eye movements in the analysis of semantically biased sentences. *Journal of Verbal Learning and Verbal Behavior*, 22, 358 –374.
- Rayner, K., Warren, T., Juhasz, B. J., & Livversedge, S. P. (2004). The effect of plausibility on eye movements in reading. *Journal of Experimental Psychology: Learning Memory and Cognition*, 30(6), 1290-1301.
- Reichle, E. D., Rayner, K., & Pollatsek, A. (2003). The EZ Reader model of eye-movement control in reading: Comparisons to other models. *Behavioral and Brain Sciences*, 26(04), 445-476.
- Saldaña, D., & Frith, U. (2007). Do readers with autism make bridging inferences from world knowledge? *Journal of Experimental Child Psychology*, 96, 310-319.
- Saldaña, D., Carreiras, M., & Frith, U. (2009). Orthographic and Phonological Pathways in Hyperlexic Readers With Autism Spectrum Disorders. *Developmental Neuropsychology*, 34(3), 240-253.
- Sansosti, F. J., Was, C., Rawson, K. A., & Remaklus, B. L. (2013). Eye movements during processing of text requiring bridging inferences in adolescents with higher functioning autism spectrum disorders: A preliminary investigation. *Research in Autism Spectrum Disorders*, 7(12), 1535-1542.



- Semel, E., Wiig, E. H., & Secord, W. A. (2003). *Clinical evaluation of language fundamentals*, fourth edition (CELF-4). Toronto, Canada: The Psychological Corporation/A Harcourt Assessment Company.
- Shah, A., & Frith, U. (1983). An insight of ability in autistic children - A research note. *Journal of Child Psychology and Psychiatry*, 24(4), 613-620.
- Snowling, M., & Frith, U. (1986). Comprehension in hyperlexic readers. *Journal of Experimental Child Psychology*, 42(3), 392-415.
- Snowling, M.J., Stothard, S.E., Clarke, P., Bowyer-Crane, C., Harrington, A., Truelove, E., Nation, K. & Hulme, C. (2010). *York Assessment of Reading for Comprehension: Passage Reading - Secondary Version*. GL Assessment.
- Tager-Flusberg, H. (1981). On the nature of linguistic functioning in early infantile Autism. *Journal of Autism and Developmental Disorders*, 11(1), 45-56.
- Taraban, R., & McClelland, J. L. (1988). Constituent attachment and thematic role assignment in sentence processing: Influences of content based expectations. *Journal of Memory and Language*, 27, 597-632 .
- Venables, W. N., & Ripley, B. D. (2002). Modern applied statistics with S (4th ed.). New York, NY: Springer. <http://dx.doi.org/10.1007/978-0-387-21706-2>
- Warren, T., & McConnell, K. (2007). Investigating effects of selectional restriction violations and plausibility violation severity on eye-movements in reading. *Psychonomic Bulletin & Review*, 14(4), 770-775.
- Warren, T., McConnell, K., & Rayner, K. (2008). Effects of context on eye movements when reading about possible and impossible events. *Journal of Experimental Psychology-Learning Memory and Cognition*, 34(4), 1001-1010.
- Wechsler, D. (1999). Wechsler Abbreviated Scale of Intelligence. The Psychological Corporation: Harcourt Brace & Company. New York, NY.

- White, S. J. (2013). The triple I hypothesis: taking another ('s) perspective on executive dysfunction in autism. *Journal of Autism and Developmental Disorders*, 43(1), 114-121.
- White, S., Frith, U., Milne, E., Rosen, S., Swettenham, J., & Ramus, F. (2006). A double dissociation between sensorimotor impairments and reading disability: A comparison of autistic and dyslexic children. *Cognitive Neuropsychology*, 23(5), 748-761.
- Williams, D. L., Goldstein, G., & Minshew, N. J. (2006). Neuropsychologic functioning in children with autism: Further evidence for disordered complex information-processing. *Child Neuropsychology*, 12(4-5), 279-298.
- Zwaan, R. A., Langston, M. C., & Graesser, A. C. (1995). The construction of situation models in narrative comprehension – An event-indexing model. *Psychological Science*, 6(5), 292-297.

Table. 1

*Means (standard deviations) for ASD and TD group's age, self reported autistic traits, intelligence, expressive language and reading skill.*

Measure	ASD			TD		
	<i>M</i>	<i>SD</i>	<i>Range</i>	<i>M</i>	<i>SD</i>	<i>Range</i>
Age	35.05	11.66	19-51	28.58	9.04	19-52
Autistic Quotient	36.86	7.15	17-49	16.79	8.12	7-35
Full scale IQ	119.50	11.18	91-140	117.54	10.14	96-139
Verbal IQ	116.73	11.31	96-137	116.00	9.42	97-138
Performance IQ	119.09	11.90	85-134	115.25	11.16	88-132
Expressive Language Raw (Max Score 96)	88.59	5.44	77-96	89.46	5.49	77-95
Single word Reading Raw (Max Score 70)	68.55	2.48	61-70	67.92	2.48	60-70
Passage Comprehension Raw (Max Score 13)	9.02	1.71	6-12	9.69	1.54	7-12

Table 2.

*An example of a plausible, implausible and anomalous sentence with region of interest boundaries marked.*

Sentence type	Regions of Interest				
	Start	Pre target	Target	Post target	Final
Plausible	The waiter used a jug to pour	the fresh	milk	in the	teacup.
Implausible	The waiter used a bucket to pour	the fresh	milk	in the	teacup.
Anomalous	The waiter used a seed to grow	the fresh	milk	in the	teacup.

Table 3.

*Global eye movement measure means (standard deviations).*

Condition	Mean Fixation Duration (ms)		Mean Fixation Count		Mean Sentence Reading Time (ms)	
	TD	ASD	TD	ASD	TD	ASD
Plausible	215 (30)	227 (42)	12 (4)	15 (7)	3122 (1037)	4245 (2161)
Implausible	217 (27)	225 (38)	12 (4)	16 (7)	3151 (1033)	4496 (2262)
Anomalous	216 (28)	228 (38)	15 (6)	18 (9)	3609 (1374)	5051 (2680)

Table 4.

*Means (standard deviations) of the eye movement measures (ms) for the pre-target, target and post target regions.*

Condition	First Fixation Duration		Single Fixation Duration		Gaze Duration		Go Past Time		Total Time	
	TD	ASD	TD	ASD	TD	ASD	TD	ASD	TD	ASD
Pre-target										
Plausible	209 (52)	221 (66)	227 (55)	221 (56)	294 (114)	325 (160)	316 (134)	352 (187)	368 (177)	524 (337)
Implausible	208 (53)	218 (67)	230 (47)	232 (74)	298 (115)	316 (158)	351 (181)	362 (207)	397 (209)	517 (370)
Anomalous	207 (53)	213 (61)	232 (55)	227 (70)	318 (130)	325 (158)	372 (196)	380 (195)	545 (288)	741 (442)
Target										
Plausible	205 (54)	233 (80)	198 (50)	236 (75)	241(99)	247 (93)	267 (136)	302 (183)	312 (182)	388 (243)
Implausible	218 (57)	225 (69)	224 (57)	232 (75)	261(92)	251 (99)	295 (142)	300 (157)	312 (144)	393 (239)
Anomalous	226 (67)	245 (86)	232 (64)	261 (90)	266 (106)	281 (104)	337 (207)	361 (205)	396 (221)	461 (258)
Post Target										
Plausible	229 (85)	239 (90)	253 (102)	256 (99)	314 (160)	332 (163)	409 (251)	578 (521)	415 (222)	509 (296)
Implausible	239 (76)	236 (80)	262 (92)	263 (84)	338 (164)	309 (144)	448 (281)	576 (584)	437 (223)	494 (274)
Anomalous	229 (79)	247 (91)	240 (79)	258 (97)	333 (164)	361 (187)	583 (494)	765 (800)	483 (257)	647 (403)

Table 5.

*Means (standard deviations) for the three measures analysed as part of the supplementary analyses.*

	Proportion of Rereading		Rereading Duration (ms)		Proportion of First Pass Regressions Out	
	TD	ASD	TD	ASD	TD	ASD
Pre-target						
Plausible	.26 (.44)	.45 (.50)	295 (159)	465 (383)	.06 (.23)	.06 (.24)
Implausible	.31 (.46)	.41 (.49)	325 (189)	589 (579)	.08 (.27)	.10 (.30)
Anomalous	.58 (.49)	.73 (.44)	412 (237)	571 (432)	.10 (.30)	.14 (.35)
Target						
Plausible	.24 (.43)	.41 (.49)	277 (141)	364 (266)	.11 (.32)	.12 (.33)
Implausible	.21 (.41)	.40 (.49)	280 (150)	380 (287)	.11 (.32)	.14 (.34)
Anomalous	.39 (.49)	.53 (.50)	324 (186)	345 (236)	.17 (.38)	.19 (.40)
Post target						
Plausible	.32 (.47)	.41 (.49)	309 (190)	428 (283)	.23 (.42)	.30 (.46)
Implausible	.30 (.46)	.45 (.50)	347 (198)	418 (276)	.23 (.42)	.31 (.46)
Anomalous	.44 (.50)	.55 (.50)	363 (227)	546 (400)	.35 (.48)	.40 (.49)

Figure Caption Sheet

*Figure 1.* Mean first fixation duration on the target word for plausible, implausible and anomalous sentences. Error bars represent standard error.



